# Identification and Characterization of Phytochemicals Regarding Their Inhibitory Potential for the Key Regulators of Cancerous Cells using Computational Biology Approaches.

**Research protocol**

for registration in the Ph.D. program

Submitted by:

**Sanjay Purohit**

Under the Guidance of

**Prof. (Dr.) Widhi Dubey**

Department of Botany

JECRC University, Jaipur

December 2021

# Contents


| Particulars | Page No. |
|---|---|
| 1. **Introduction** | **1-2** |
| 2. **Review of the literature** | **2-4** |
| 3. **Statement of the problem** | **4** |
| 4. **Objective of the study** | **5** |
| 5. **Limitations** | **5** |
| 6. **Significance of the study** | **5** |
| 7. **Methodology** | **6-7** |
| 8. **References** | **8-9** |

Dated: …….

# Certificate

This is to certify that the research protocol entitled "**Identification and Characterization of Phytochemicals Regarding their Inhibitory Potential for the Key Regulators of Cancerous Cells using Computational Biology Approaches"** by **Sanjay Purohit**, submitted in fulfillment of the requirement for the registration in the Ph.D. program in the Department of Botany under faculty of Science at JECRC University, Jaipur, during the academic year 2021-2022, is a bonafide record work carried out under our guidance and supervision.

**Prof. (Dr.) Widhi Dubey**

Department of Botany

JECRC University, Jaipur

director.science@jecrcu.edu.in

**Introduction:**

Cancer, a large group of diseases arises due to abnormal and uncontrollable growth and division of the cells, leading to the formation of cell mass or tumor. This uncontrollable growth and division of the cells can start in almost any tissue or organ of the body in virtually any stage of the life span. Cancer cells not only show indefinite growth in the organ which they initially develop but they go beyond the usual boundaries to invade adjoining tissue and even spread to other organs. The invasion of cancer cells to other body parts is called metastasis, is a major cause of death due to cancer as a disease. Cancerous tumors are also called neoplastic or malignant tumors.

At a global level, cancer becomes the second leading cause of death annually. In the year 2018, the estimated number of deaths caused by cancer was 9.6 million which was almost one in six deaths, occurred in that year. Lung, prostate, breast, colorectal, stomach, thyroid, cervical, liver cancers, etc. are some of the most common types of cancer that occur in human beings. The occurrence of these or other types of cancers varies according to gender or age groups.

In the last 50 years, there has been a huge increase in the global cancer burden, and it is continuously rising. This has exerted tremendous emotional, physical, and financial strain on individuals, families, societies, and the overall healthcare system of countries. The developed countries like the USA, UK, Germany, etc. where the healthcare system is strong, patients have early detections of cancerous conditions and also have access to quality treatment. Thus, the survival rates of the patients, in those countries for many types of cancers are relatively good. Even though in 1985, Dr. John Bailer declared the USA's national cancer program as a "qualified failure" in terms of the number of total incidents of cancer cases and the total number of patients who died due to cancer.

So, in terms of the emergence of cancer cases, the situation is alarming for the countries with strong healthcare and economical structure, but, on the contrary, the healthcare system in low- and middle-level economic stature countries like India is least equipped to manage this burden, which makes the condition disastrous. In these countries, most cancer patients do not get the appropriate diagnosis or quality treatment.

As studies suggested the development of cancer the role of genetic or hereditary defects is rare, while it mainly depends upon the lifestyle, environment & dietary habits of individuals. So, by transformation into a healthy lifestyle like avoiding tobacco, environmental pollutants, calorie restrictions can delay the progression of cancer in many individuals. Studies have also

suggested that certain phytochemicals (i.e., plant chemicals), available in regular fruit and vegetable diets or available in different medicinal plants have anti-cancerous activities. for countries like India, where healthcare is not so strong, treatment of cancer with the help of dietary phytochemical or medicinal phytochemicals could be a good remedy for the bulk of the patients affected with this disease.

Phytochemicals are chemical compounds produced by different parts of the plant such as fruits, seeds, vegetables, etc. The role of certain phytochemicals such as vitamins, terpenes, alkaloids, etc. in the major metabolic pathways and biosynthesis of different molecules of our body has been intensively studied and widely known, yet only for a handful of such phytochemicals like curcumin, EGCG, etc. their detailed regulatory function in different signal transduction processes is known. Over the last decade, several studies have reported regulation of cell proliferation pathways, microRNAs & cancer stem cell maintenance via different phytochemicals. So, it can easily be concluded that there could be many of such phytochemicals whose function in the regulation of key cell signaling molecules, hence in the treatment of cancer, is still unknown to us. Formation of a curated database of phytochemicals of Indian medicinal plants (IMPPAT; Indian Medicinal Plants, Phytochemistry And Therapeutics) in 2018 (Mohanraj et al., 2018), now the possibility of finding such medicinally relevant phytochemicals has increased enormously.

**Review of the literature:**

In the last couple of decades, a lot of natural phytochemicals and dietary chemicals were studied for their anti-cancerous activity. Curcumin, a kind of polyphenol, obtained from *Curcuma* species is one of the most studied phytochemicals for its antioxidant and tumoricidal properties against different kinds of cells. Rats having AK-5 tumor cells are treated with curcumin show induction of DNA fragmentation, loss of membrane potential of mitochondria, and induction of superoxide radicals which eventually lead to caspase-mediated-apoptotic cell death of the tumor cells (Bhaumik et al., 1998). In older human beings, dietary curcumin also prevents obesity-associated vascular dysfunction (Kuszewski et al., 2020).

Cancer cells evade the cell cycle regulations and hence show unchecked & accelerated cell division. Cell cycle regulation is usually carried out by cyclin proteins and cyclin-dependent kinases (CDKs). Cyclin $D_1$ is one such cyclin whose activity is crucial for a cell in $G_1$/S phase transition in the cell cycle. Cyclin $D_1$ activates CDK4 and CDK6, and these kinases

phosphorylate retinoblastoma (RB) protein and the tumor cell surpasses the $G_1$-cell arrest (Matsushime et al., 1994). Hence Cyclin $D_1$ is overexpressed in most of the cancer cell lines. Various breast cancer cell lines e.g., MCF-7 and prostate cancer cell line, LNCaP were treated with curcumin to show the inhibition of the expression of cyclin $D_1$ by its proteolysis, which leads to inhibition of proliferation of these cancer cell lines (Mukhopadhyay et al., 2002).

Nuclear Factor kappa B (NF-κB) is an important transcription factor that regulates numerous cell-signaling pathways. NF-κB upregulates the various proinflammatory cytokines like Interleukin (IL)-1β, IL-6, IL-12, interferon (IFN)-γ and tumor necrosis factor (TNF)-α (Rogler et al., 1998). These cytokines play important role in colonic inflammation, ulcerative colitis, colon cancers, and myeloma. Curcumin downregulates the level of NF-κB to suppress the level of these pro-inflammatory cytokines and inflammation-mediated progression of cancer in colon cells (Singla, 2013). Studies also suggest that IL-6 stimulates phosphorylation of STAT3, a cell-signaling protein, and thus stimulates the cell proliferation of multiple myeloma cells. Curcumin inhibits IL-6-induced STAT3 phosphorylation and its nuclear transfer to eventually inhibit cell proliferation (Bharti et al., 2003).

Resveratrol, another kind of polyphenol, is abundant in many plant species and is also a part of the human diet, especially in the form of grapes and red wine. Similar to curcumin, resveratrol also shows anti-oxidant and anti-inflammatory properties. Resveratrol prevented the development of preneoplastic lesions in rat mammary gland cells treated with a carcinogen in culture and also prevented tumorigenesis in a mouse skin cancer model (Jang, 1997). Resveratrol-enriched diet in mice also reduces the level of pro-inflammatory cytokines like TNF-α and IL-1β significantly (Sanchez-Fidalgo et al., 2010). Due to its anti-inflammatory properties, resveratrol can prevent chemical-induced neurotoxicity (Tang et al., 2021).

Resveratrol has a potential therapeutic effect via the promotion of immune surveillance through the innate immune system thereby eliminating the development of spontaneous tumor cells before their proliferation. Natural Killer (NK) cells are the primary effector cells of our system which can identify significantly transformed or infected cells without any prior antigen presentation to them. Because of this characteristic of NK cells, they can effectively eliminate many fast-growing tumor cells before proliferation (Bui et al., 2007). NKG2D is the antigen receptor that plays a crucial role in tumor surveillance. It is expressed by CD8 T cells and NK cells which use the NKG2D receptor to recognize specific surface ligands. These ligands are

expressed at a very low level in normal cells but are overexpressed by tumors and some infected cells (Guerra et al., 2008).

Epigallocatechin-3-gallate (EGCG), a catechin polyphenol, abundantly found in green tea and black teas, is a strong antioxidant compound and as an antioxidant, it prevents DNA damage caused by reactive oxygen species and hence prevents the mutation of normal cells which can convert them into tumors (Seeram et al., 2006). Pre-clinical studies confirm that EGCG inhibits tumor cell growth by inducing apoptosis, thus preventing cancer development (Yiannakopoulou, 2014).

Dietary phytochemicals are well known for their antioxidant or chemopreventive properties, but they are indeed very effective epigenetic regulators as well, which may target or even revert abnormal epigenetic modifications in various human pathological conditions, including cancer. These dietary phytochemicals have shown potential as DNA methylation inhibitors, and histone modifiers, and they can also restore gene expression of tumor-suppressor genes altering expression levels of various miRNAs (Shankar et al., 2016).

Natural dietary products have such an important role as new sources of medicines. However, in the last few decades, there has been a huge decrease in the number of marketable drugs that are derived from natural sources. One crucial reason behind that was the unavailability of a comprehensive database of different plants having medicinal properties and the chemicals produced by those plants but due to the recent development of Indian Medicinal Plants, Phytochemistry And Therapeutics (IMPPAT), a well-curated & openly accessible database, now it is possible to look in the unexplored sources of natural products to be used as potential drugs (Mohanraj et al., 2018).

**Statement of the problem:**

**Identification and characterization of phytochemicals regarding their inhibitory potential for the key regulators of cancerous cells using computational biology approaches.**

**Objectives of the study:**

- To identify common cell signaling intermediates that promote or allow sustenance of cancer conditions in different types of cancers using data curing and high throughput RNA sequencing data analysis.
- To identify and characterize phytochemical inhibitors like polyphenols, catechins, lycopene etc. against cell signaling intermediates common in various types of cancer using *in-silico* molecular docking studies.
- To identify and characterize phytochemical inhibitors polyphenols, catechins, lycopene etc. against well-studied miRNA and genetic regulators, important in the progression of different cancers using *in-silico* molecular docking studies.

**Limitations:**

- This study will include computation identification and characterization of novel phytochemicals having which could be used as potential drugs to cure certain cancer. Since being a computational study, it needs to be validated or confirmed by wet-lab experiments.
- This study will not take into account the delivery system to be used to deliver the potential phytochemical drug in a patient.
- Since there are millions of plant chemicals available in any given region, this study will be associated with a few major groups of phytochemicals.

**Significance of the study:**

- As the study is about the exploration of unexplored plant chemicals, it has good potential to provide references of novel chemicals which could be used as drugs in the future.
- Identification of common regulators or common nodes could simplify our approach for a cure for complex diseases like cancer, which has numerous types and triggers of its promotion.

**Methodology:**

1. **High-throughput RNA sequencing (RNA-seq) for transcriptomic studies and selection of target molecules:** This technique uses next-generation sequencing methods to quantify the whole transcriptome of a cell or group of cells. It is popularly known as **RNA-seq**. Since the last decade, RNA-seq has become a vital tool in the field of molecular biology because of its ability to identify post-transcriptional modifications, alternative gene spliced transcripts, SNPs, or changes in gene expression over time, or differences in gene expression in different cell types or treatment groups.

   By analysing the differential expression of cell signaling intermediate proteins, enzymes, and miRNAs in cancer cell lines/tissues in comparison to the non-cancer/ normal human cells presents a great opportunity to identify common key regulators across different types of cancers that help in the maintenance of cancer condition.

   Raw data obtained by high-throughput RNA sequencing experiments done by different research groups in various types of cancer cell lines, in different conditions/treatment is available on various open-access databases such as NCBI-SRA, EBI, etc. which is accessible to all. Different approaches or pipelines could be used to analyse this data. One such pipeline is "**Cufflinks-suite**" which is maintained by Cole Trapnell's lab at the University of Washington. It estimates the abundance and differential expression of transcripts across different treatment groups, such as cancer vs non-cancer or cancer cells treated with certain drugs vs cancer cells without any treatment, etc.

   The detailed methodology or software used in the cufflinks-suite are as follow:

- **Bowtie2:** Alignment of raw RNA-seq reads against reference human or another genome for identification of the RNA reads. It produces output in the form of SAM (Sequence Alignment Map) or BAM (Binary Alignment Map) files.
- **TopHat:** Similar to Bowtie2, TopHat is also involved in the alignment of RNA sequence reads to the reference genome but its most important significance is that it can align those reads which Bowtie2 cannot align. is used to identify the splice variants generated due to alternative splicing. The output of TopHat is in the form of BAM files.
- **Cufflinks:** It counts the RNA sequenced reads, aligned to a specific region of the reference genome in a particular sample. It provides the absolute number of reads corresponding to

each gene or DNA region along with the statistical significance of aligning the reads in the form of the p-value. The output of cufflinks is obtained in the form of notepad or MS-Excel files.

- **Cuffdiff:** It provides the differential count of RNA sequencing reads corresponding to each gene or DNA region across different samples. Cuffdiff normalizes the variability among different samples and shows comparative sequence reads across the samples. With the output of cuffdiff, upregulation or downregulation of any gene in a specific sample/condition can easily be estimated.
- **Trimmomatic:** It is an additional or optional method to trim off the ambiguous, low-quality sequence reads or the adaptor used in the experiment kit for the preparation of RNA samples for sequencing.

2. **Selection of Novel phytochemical having medicinal potential:** It will be done using manual or coding-based curation of relevant data available at IMPPAT; Indian Medicinal Plants, Phytochemistry And Therapeutics database.

3. **Structural analysis of target molecule and phytochemical structure:** To study the interaction between a potential target molecule and a phytochemical, firstly the molecular structures of these compounds have to be obtained or prepared. Most of these structures can be obtained from the "**Protein Data Bank (PDB)"** and "**PubChem",** a chemical data base. These are open access databases from which the crystal structures of different compounds and their possible isoforms can be obtained.

   For further interaction studies between possible target and phytochemical, the following methods or software to be used:

- **I-tasser (Iterative Threading Assembly Refinement) & Swiss-model:** These are server-based approaches for protein structure prediction and structure-based function annotation. These are useful for those proteins whose crystal structures are not available in PDB.
- **Autodock & Autodock Vina:** These programs are widely used for molecular docking. Docking is a technique of *in-silico* (computational) prediction of the binding or interaction between two molecules. Molecular docking can predict novel interacting partners/inhibitors for specific molecules and also can predict the association efficiency of two interacting molecules.

- **PyMol & RasMol:** These are widely used molecular graphics programs for the visualization of proteins, nucleic acids, and small molecules. Using these programs, we can visualize individual molecule as well as their docked interactive structure.

**References:**


1. Bharti AC, Donato N, Aggarwal BB. (2003). Curcumin (Diferuloylmethane) Inhibits Constitutive and IL-6-Inducible STAT3 Phosphorylation in Human Multiple Myeloma Cells. Texas, USA. Journal of Immunology, 171, 3863–3871.
2. Bhaumik S, Anjum R, Rangaraj N, Pardhasaradhi BVV, Khar A. (1998). Curcumin mediated apoptosis in AK-5 tumor cells involves the production of reactive oxygen intermediates. CCMB Hyderabad, India. FEBS letters, 456, 311-314.
3. Bui JD, Schreiber RD. (2007). Cancer immunosurveillance, immunoediting and inflammation: independent or interdependent processes? University of California at San Diego, USA. Current Opinion in Immunology, 19, 203-208.
4. Guerra N, Tan Y, Joncker N, Choy A, Gallardo F, Xiong N, Knoblaugh S, Cado D, Greenberg N, Greenberg N, Raulet D. (2008). NKG2D-deficient mice are defective in tumor surveillance in models of spontaneous malignancy. University of California at Berkeley & Fred Hutchinson Cancer Research Center, Seattle, USA. Immunity, 28, 571-580.
5. Jang M, Cai L, Udeani GO, Slowing KV, Thomas CF, Beecher CW, Fong HH, Farnsworth NR, Kinghorn AD, Mehta RG, Moon RC, Pezzuto JM. (1997). Cancer chemopreventive activity of resveratrol, a natural product derived from grapes. University of Illinois at Chicago, USA. Science, 275, 218–220.
6. Kuszewski JC, Howe PRC, Wong RHX. (2020). Evaluation of Cognitive Performance following Fish-Oil and Curcumin Supplementation in Middle-Aged and Older Adults with Overweight or Obesity. Australia. Journal of Nutrition, 10, 150(12), 3190-3199.
7. Matsushime H, Quelle DE, Shurtleff SA, Shibuya M, Sherr CJ, Kato JY. (1994). D-type cyclin-dependent kinase activity in mammalian cells. University of Tokyo, Japan. Molecular and Cellular Biology, 14, 2066-2076.
8. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, Samal A. (2018). IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. Chennai, India. Scientific Reports, 8, 4329.

9. Mukhopadhyay A, Banerjee S, Stafford LJ, Xia C, Liu M, Aggarwal BB. (2002). Curcumin induced suppression of cell proliferation correlates with down-regulation of cyclin D1 expression and CDK4-mediated retinoblastoma protein phosphorylation. Texas, USA. Oncogene, 21, 8852–8862.
10. Rogler G, Andus T. (1998). Cytokines in inflammatory bowel disease. University of Regensburg, Germany. World Journal of Surgery, 22, 382–389.
11. Sanchez-Fidalgo S, Cardeno A, Villegas I, Talero E, De la Lastra CA, (2010). Dietary supplementation of resveratrol attenuates chronic colonic inflammation in mice. University of Seville, Spain. European Journal of Pharmacology, 633, 78–84.
12. Seeram NP, Henning SM, Niu Y, Lee R, Scheuller HS, Heber D. (2006). Catechin and caffeine content of green tea dietary supplements and correlation with antioxidant capacity. University of California, Los Angeles, USA. Journal of Agricultural and Food Chemistry, 54, 1599-1603.
13. Shankar E, Kanwal R, Candamo M, Gupta S. (2016). Dietary phytochemicals as epigenetic modifiers in cancer: Promise and challenges. Ohio, USA. Seminars in Cancer Biology, 40-41, 82–99.
14. Singla V, Mouli VP, Garg SK, Rai T, Choudhury BN, Verma P, Deb R, Tiwari V, Rohatgi S, Dhingra R, Kedia S, Sharma PK, Makharia G, Ahuja V. (2013). Induction with NCB-02 (curcumin) enema for mild-to-moderate distal ulcerative colitis—A randomized, placebo-controlled, pilot study. AIIMS Delhi, India. Journal of Crohn's and Colitis, 8, 208–214.
15. Tang XL, Wang X, Fang G, Zhao YL, Yan J, Zhou Z, Sun R, Luo AL, Li SY. (2021). Resveratrol ameliorates sevoflurane-induced cognitive impairment by activating the SIRT1/NF-κB pathway in neonatal mice. Huazhong University of Science and Technology, Wuhan, China. Journal of Nutritional Biochemistry, 2021, 90, 108579.
16. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Barbara J Wold BJ, Pachter L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. USA. Nature Biotechnology, 28, 5.
17. Yiannakopoulou EC. (2014). Effect of green tea catechins on breast carcinogenesis: a systematic review of in-vitro and in-vivo experimental studies. Technological Educational Institute of Athens, Greece. European Journal of Cancer Prevention, 23, 84-89.